Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Robust brain extraction tool for CT head images

Zeynettin Akkus*, Petro Kostandy, Kenneth A. Philbrick, Bradley J. Erickson

Radiology Informatics Lab, Department of Radiology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA

ARTICLE INFO

Article history:
Received 28 May 2018
Revised 20 November 2018
Accepted 9 December 2018
Available online xxx

Keywords:
Convolutional neural network
Deep learning
Image segmentation
Skull stripping
Brain extraction
Computed tomography

ABSTRACT

Extracting brain parenchyma from computed tomography (CT) images of the head is an important prerequisite step in a number of image processing applications, as it improves the computational speed and accuracy of quantitative analyses and image co-registration. In this study, we present a robust method based on fully convolutional neural networks (CNN) to remove non-brain tissues from head CT scans in a computationally efficient manner. The method includes an encoding part, which has sequential convolutional filters that produce feature representation of the input image in low dimensional space, and a decoding part, which consists of convolutional filters that reconstruct the input image from the reduced representation. We trained several CNN models on 122 volumetric head CT scans and tested our models on 22 withheld volumetric CT head scans based on two experts' manual brain segmentation. The performance of our best CNN model on the test set is: Dice Coefficient = 0.998 ± 0.001 (mean ± standard deviation), recall = 0.999 ± 0.001, precision = 0.998 ± 0.001, and accuracy = 1. Our method extracts complete volumetric brain from head CT images in about 2 s which is substantially faster than currently available methods. To the best of our knowledge, this is the first study using CNN to perform brain extraction from CT images. In conclusion, the proposed approach based on CNN provides accurate extraction of brain tissue from head CT images in a computationally efficient manner.

1. Introduction

Computed tomography (CT) imaging of the head is a very commonly performed procedure, which provides detailed information about brain tissue, structures and pathology of the brain. Image reconstructions are typically obtained using both a soft tissue and bone kernels, which allow optimal evaluation of the brain parenchyma and osseous structures respectively. The bone kernel retains higher frequency image information, resulting in images that are both sharper and noisier. Noncontrast CT is usually the first-line imaging modality for evaluating patients presenting to the emergency room with a multitude of neurological complaints. It is very accessible, images are rapidly acquired, often in conjunction with other body parts, and there are fewer contraindications than MRI. Early detection of hemorrhage and stroke from noncontrast head CT is particularly important for guiding appropriate management [18]. Several image processing methods using CT scans have been developed to detect hemorrhage [13,25,26] and acute infarcts [6,18,19], segment brain tissue and structures [29], register intra- and inter- modality images [9,18,27], and to perform surface reconstruction. Removing non-brain tissues such as the skull, scalp, and face from head CT images as well as inanimate objects such as the table, head holder and external medical devices is a prerequisite step in many of these analyses. It is desirable to automate this task, both to reduce human variability and eliminate time-consuming manual segmentation that would be infeasible in large-scale clinical studies.

Brain extraction has been extensively studied for magnetic resonance (MR) imaging of the brain [5,10,24,28] but not for head CT images. Bauer et al. [5] presented a brain extraction algorithm based on Insight Toolkit (ITK) [23] for both MR and CT images. Their method involves using atlas-based registration to create a rough mask of the brain parenchyma, which is subsequently used to initialize the level set algorithm for a more refined end result. This method is included as part of ITK and is publicly available for use. Brain Extraction Tool (BET) [24] is another popular tool for extracting brain tissue from MR images of the head but it does not work on CT head images. Muschelli et al. [16] modified the parameters of the BET algorithm to make it suitable for CT. Their method includes thresholding each image to brain tissue range (0–100 HU), smoothing the thresholded image with a 3D Gaussian kernel, and modifying the Fractional Intensity (FI) parameter of BET. The FI parameter varies between 0 and 1, and determines the edge of the segmented brain image. The smaller FI results in the larger brain masks. The CT BET algorithm is publicly available. Similar to

* Corresponding author.
E-mail addresses:
akkus.zeynettin@mayo.edu (Z. Akkus), kostandy.petro@mayo.edu (P. Kostandy), philbrick.kenneth@mayo.edu (K.A. Philbrick), bje@mayo.edu (B.J. Erickson).

https://doi.org/10.1016/j.neucom.2018.12.085
0925-2312/

Muschelli et al. [16], Rordenet al. [21], Roden and Brett [22], and Yang et al. [30] adapted BET for brain extraction but they did not validate their approach. The CT attenuation of tissues such as eyes and neck, which is similar to that of brain tissue, make threshold based approaches less effective in these regions. Another study intended to extract brain from CT head images uses a particle filter based segmentation approach that was introduced by Mandell et al. [14]. They used a limited dataset however,five pediatric patient head CT scans, to validate their method.

In this study, we present a fully automated approach based on a fully convolutional neural networks (CNN) for extracting brain tissue from head CT images in a few seconds, which outperforms previous approaches. To the best of our knowledge, this is the first study using CNN's for extracting brain tissue from CT head images. We made our best model and its training weights publicly available on github (https://github.com/aqqush/CT_BET.git) for the use of other researchers.

2. Materials and methods

2.1. Data acquisition and processing

We retrospectively collected 147 patients that have non-contrast head CT and MR scans for routine clinical care for stroke acquired between 2009 and 2012 at Mayo Clinic. The time of stroke onset was based on clinical history and entered into a stroke registry by an independent observer. Acute stroke was radiographically confirmed by an experienced neuroradiologist for these patients. Three cases were excluded due to the presence of major post-operative changes that include craniotomy. We did not exclude any CT head scans due to beam hardening and motion artifacts but we also did not observe any significant motion artifacts on 147 head CT images. The CT scans had dimensions of $512 \times 512 \times$ N, corresponding to the number of individual voxels in the x, y, and z planes respectively, and were acquired using a CT Siemens Sensation 64 with the following scanning parameters: 120 kV, 570–590mAs (tube current), and $0.488 \times 0.488 \times 5$ mm voxels. The mean and standard deviation of the slice number of CT head scans were 33 and 2, respectively. We did not apply any preprocessing to our images as CT images provide universally dimensionless Haunsfield Unit (HU) which is standardized for CT scans and typically ranges from -1000 to $+1000$, air vs. bone. The brain tissue was volumetrically segmented in each case using ITK-SNAP [31] by a radiologist (Expert 1). We used an HP Z820 workstation with Intel Xeon E5-2640 2.5 GHz processor and an NVIDIA Titan X Graphical Processing Unit (GPU) card that has 12GB memory for processing the data. CNN models were implemented with Keras deep learning package [7] using the Tensorflow [1] backend.

2.2. Robust brain extraction tool (rBET)

A typical fully CNN architecture contains encoding and decoding parts that include multiple layers of convolution, maxpooling, dropout, and activation. Convolutional layers produce feature maps by convolving a convolutional kernel across the input image. Maxpooling is used to downsample the output of this convolution by using the maximum value of the defined neighborhood as the value passed to the next layer. Rectified Linear Unit (ReLU) is one of the most commonly used activation functions. It nonlinearly transforms data by clipping any negative input values to "0" while positive input values are passed as output. Dropout is a regularization technique that randomly deactivates a fraction of neurons in a layer to prevent overfitting. To perform a prediction from input data, the output scores of final CNN layer are connected to a softmax nonlinearity function that normalizes scores into multinomial distribution over labels. Also, a loss function that minimizes the error between prediction and ground truth labels and optimization function (e.g. stochastic gradient descent-SGD) that updates weights at each iteration is used to train CNN architectures until it converges a steady state [2].

We used five encoder-decoder type CNN models that map the input image to the output mask for pixel-wise classification of brain and non-brain tissues in CT images. We used 2D U-Net [20], two modified versions of 2D U-Net, 3D U-Net, and SegNet [4] CNN architectures for this purpose. Fig. 1 shows a schematic illustration of a typical encoder-decoder CNN. CNN model 1 which is exactly the same to U-Net architecture contains an encoding part that has 10 convolutional layers using 3×3 kernels and rectified linear unit (ReLU) activation function, 4 max pooling layers to produce activation maps of the input image in low dimensional space, and a decoding part with 9 convolutional layers of 3×3 kernels and 4 deconvolutional upsampling layers that reconstructs the output mask from the reduced low dimensional representation of the input image at original resolution (see Fig. 2 for the model parameters). The number of feature maps and the order of layers are the same as the ones used in U-Net architecture [20]. The network weights were initialized with Glorot uniform distribution [8]. We used softmax function at the end of the network to produce two-class (i.e. brain and non-brain) classification of the output. We also used dropout regularization [17] in the encoding part to prevent overfitting of the CNN. To update the weights during the training, we used Adam [12] optimizer with initial learning step of 1e-5. Weighted categorical cross entropy loss was used to minimize the error between true and prediction labels (see Eq. (1)).

$$L = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{m} y_{ij}\log\hat{y}_{ij}w_{ij}, \tag{1}$$

where n is number of samples, m is number of classes, y is true labels, $\hat{y}$ is predicted labels, and w_{ij} is weighting for each sample of classes. $w_{ij} = \max_i\{n_0, n_1, \ldots n_j\}/n_j$ is defined to balance the impact of each class in the loss function.

CNN model 2 is a variation of model 1 in which we replaced maxpooling layers with convolutional layers with a stride of 2 as an alternative method of down-sampling, and ReLU activation with leakyReLU activation function to capture more details (see Fig. 2). CNN model 3 is a variation of model 1 that includes batch normalization and residual connections between corresponding convolutional blocks in the encoding and decoding parts. CNN model 4 is using the SegNet architecture. The encoding part of this model includes 5 blocks of stacked convolutional layers, batch normalization, ReLU activation function, and a maxpooling layer. The decoder part includes 5 blocks of stacked upsampling layer, convolutional layers, batch normalization, and RELU activation. This model uses upsampling layer in decoder rather than deconvolutional layer used U-Net. Model 5 is the 3D Unet architecture which is trained on four times downsampled CT head images due to memory limitation on GPU cards. The predicted labels in the 3D-UNet was upsampled four-times to match the original input image size.

2.3. Experiments and evaluation

We divided our data into 122 scans for training and 22 scans for testing. We applied 5-fold cross validation to the training data to assess the generalization ability of CNN model 1. CNN models 1–5 were trained on entire training dataset and compared based on their performance on the test dataset. The test dataset was segmented by two experts (Expert 1 and 2) to assess the interobserver variability in manual segmentation. Only the manual segmentations of Expert 1 are used as the ground truth for the training phase.

Please cite this article as: Z. Akkus, P. Kostandy and K.A. Philbrick et al., Robust brain extraction tool for CT head images, Neurocomputing, https://doi.org/10.1016/j.neucom.2018.12.085

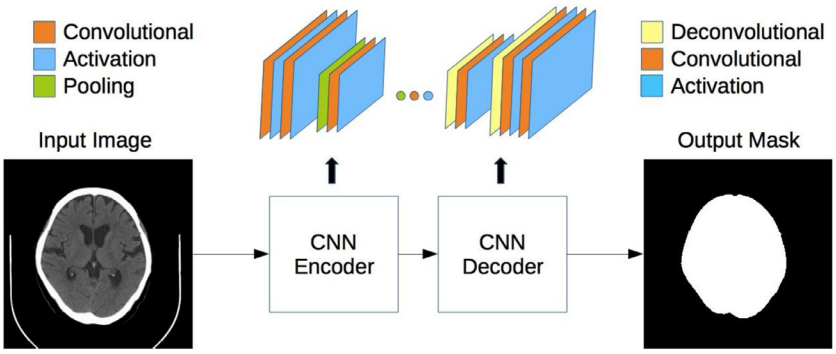

Fig. 1. Schematic illustration of an encoder-decoder type convolutional neural network.

Layer	Output Shape	Kernel Size	Stride	Activation
Input	512×512×1			
Conv1	512×512×64	3×3	1×1	RELU
Conv2	512×512×64	3×3	1×1	RELU
MaxPooling1	256×256×64	2×2		
Conv3	256×256×128	3×3	1×1	RELU
Conv4	256×256×128	3×3	1×1	RELU
MaxPooling2	128×128×128	2×2		
Conv5	128×128×256	3×3	1×1	RELU
Conv6	128×128×256	3×3	1×1	RELU
MaxPooling3	64×64×256	2×2		
Conv7	64×64×512	3×3	1×1	RELU
Conv8	64×64×512	3×3	1×1	RELU
Dropout1 (p=0.5)	64×64×512			
MaxPooling4	32×32×512	2×2		
Conv9	32×32×1024	3×3	1×1	RELU
Conv10	32×32×1024	3×3	1×1	RELU
Dropout2 (p=0.5)	32×32×1024			

Layer	Output Shape	Kernel Size	Stride	Activation
Output	512×512×2			Softmax
Conv19	512×512×2	1×1	1×1	RELU
Conv18	512×512×64	3×3	1×1	RELU
Conv17	512×512×64	3×3	1×1	RELU
Deconv4+Conv2	512×512×128		Concatenation	
Deconv4	512×512×64	2×2	2×2	Linear
Conv16	256×256×128	3×3	1×1	RELU
Conv15	256×256×128	3×3	1×1	RELU
Deconv3+Conv4	256×256×256		Concatenation	
Deconv3	256×256×128	2×2	2×2	Linear
Conv14	128×128×256	3×3	1×1	RELU
Conv13	128×128×256	3×3	1×1	RELU
Deconv2+Conv6	128×128×512		Concatenation	
Deconv2	128×128×256	2×2	2×2	Linear
Conv12	64×64×512	3×3	1×1	RELU
Conv11	64×64×512	3×3	1×1	RELU
Deconv1+Dropou	64×64×1024		Concatenation	
Deconv1	64×64×512	2×2	2×2	Linear

Encoding Part
Decoding Part
Convolution Blocks
Deconv ==> Deconvolution Layer
Conv ==> Convolutional Layer
RELU ==> Rectified Linear Unit

Fig. 2. The parameters used for Unet CNN artchitecture [20]. Decon+Conv shows the concatanation of layers from the encoding and deconding parts, also called skip connections.

To increase the number of training samples and make the learning process robust to variations, we augmented the training data by applying random transformations, including in-plane rotation (±15° in x-y axial plane), horizontal flipping, translation (± 5 pixels in x and y directions), and scaling (±10% of original size). To further assess the impact of data augmentation, we collected a single head CT scan with 1 mm slice thickness and assessed the generalization ability of CNN model 1 network training only on this case with and without data augmentation. As the slice thickness for this scan is much lower than that of training data, we applied random 10 rotations around each axis (range: −90° to +90°), which would not be possible for the training data due to interpolation error caused by thick image slices. The images are resampled with cubic spline interpolation.

To determine in which epoch to stop training, we saved CNN model 1 wt in every 25 epochs and selected the epoch where the training loss stopped changing at the multiple of 25 epochs. We opted to stop training after 100 epochs, as the training loss essentially plateaus after 80 epochs, as can be seen in Fig. 3. This is observed for the each left out fold of 5-fold cross validation experiment.

To assess the impact of different loss function, we trained CNN model 1 with categorical cross entropy loss and Dice coefficient loss [15]. To assess the performance of rBET on head CT images

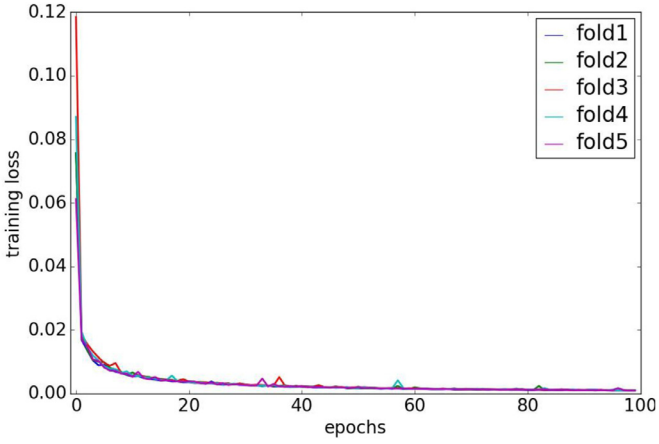

Fig. 3. Training loss curve for each left out fold of five-fold cross validation.

reconstructed with Bone kernel, we trained CNN model 1 on CT head images reconstructed with Bone and soft tissue kernels separately and together. We also visually assessed the performance of our best model on three post-surgical and two neonatal head CT scans.

Please cite this article as: Z. Akkus, P. Kostandy and K.A. Philbrick et al., Robust brain extraction tool for CT head images, Neurocomputing, https://doi.org/10.1016/j.neucom.2018.12.085

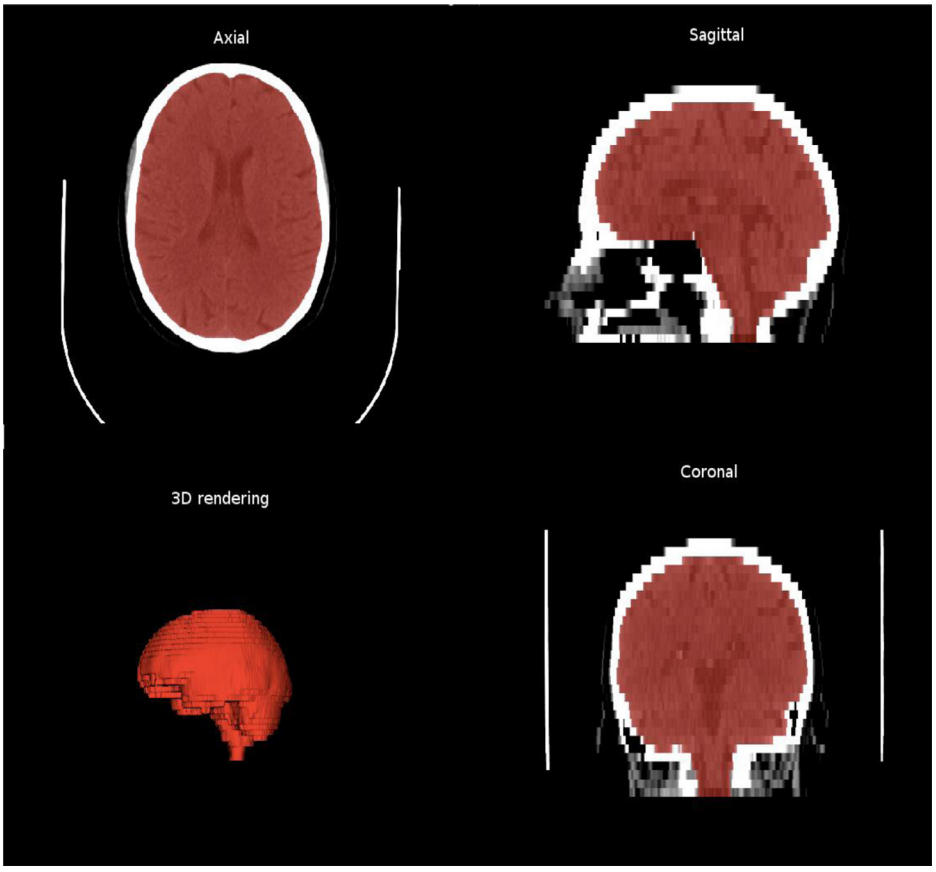

Fig. 4. An example of brain extraction from CT images using CNN model 1.

Table 1
The performance (mean ± standard deviation) of our brain extraction tool for CT head images on the test dataset. Execution time is for the segmentation of a complete volume.

Models	Dice	Accuracy	Precision	Recall	Execution time (seconds)
1 (2D)	0.998 ± 0.001	1.000 ± 0.000	0.998 ± 0.001	0.999 ± 0.001	2 s
2 (2D)	0.998 ± 0.001	0.999 ± 0.000	0.997 ± 0.001	0.999 ± 0.001	2 s
3 (2D)	0.993 ± 0.003	0.998 ± 0.001	0.987 ± 0.007	0.998 ± 0.001	2 s
4 (2D)	0.991 ± 0.001	0.998 ± 0.000	0.985 ± 0.001	0.998 ± 0.001	3 s
5 (3D)	0.990 ± 0.002	0.997 ± 0.001	0.985 ± 0.001	0.994 ± 0.004	~10 s

We compared the performance of rBET to those of Brain Extraction Tool (BET) [11] modified for CT images [16] (CT BET) and the ITK skull stripping filter (ITK Filter) on test dataset. The CT BET and ITK skull stripping filter were run on the computer processing unit (CPU). To compare the performance of our method to publicly available methods, we calculated the Dice coefficient, accuracy, precision, and sensitivity metrics [2].

We also visually evaluated the performance of our best CNN model (1) on three post-surgical scans, two neonatal scans, and fifty-two scans with extra-axial hemorrhage.

3. Results

An example of brain extraction is seen in Fig. 4. Table 1 shows the performance of the five CNN models over 22 test cases. As seen in Table 1, CNN models 1 showed the best performance. Training the CNN model 1 with categorical cross entropy and dice loss function showed the same performance on the test dataset. Replacing the maxplooling layers and using leaky ReLU in CNN model 2 slightly decreased the precision. CNN models 3, 4, and 5 showed a lower performance compared to model 1. The five-fold cross validation results in Table 2 show almost no variability between each left-out fold indicating good generalization. Table 3 and Fig. 5 show the performance of other publicly available methods. ITK skull stripping filter was the least accurate compared to other methods. Table 3 also shows the performance of the rBET trained on only one thin slice CT scan with and without data augmentation. The rBET trained with aggressive data augmentation showed improved performance level compared to the one trained without data augmentation. CT BET provided the closest results to our proposed approach in head CT images reconstructed with soft tissue kernel but it showed noticeable drop in performance when tested on head CT images reconstructed with the bone kernel as seen in Table 4. The performance of our CNN model slightly dropped but is still high on

Table 2
Five-fold cross validation results (mean ± standard deviation) on training dataset.

Model 1	Dice	Accuracy	Precision	Recall
Fold 1	0.998 ± 0.001	0.999 ± 0.000	0.997 ± 0.001	0.999 ± 0.001
Fold 2	0.998 ± 0.001	0.999 ± 0.000	0.997 ± 0.001	0.999 ± 0.001
Fold 3	0.998 ± 0.001	1.000 ± 0.000	0.998 ± 0.001	0.999 ± 0.001
Fold 4	0.998 ± 0.000	1.000 ± 0.000	0.997 ± 0.001	0.999 ± 0.000
Fold 5	0.998 ± 0.001	1.000 ± 0.000	0.997 ± 0.001	0.999 ± 0.000

Please cite this article as: Z. Akkus, P. Kostandy and K.A. Philbrick et al., Robust brain extraction tool for CT head images, Neurocomputing, https://doi.org/10.1016/j.neucom.2018.12.085

Table 3
The performance (mean ± standard deviation) of the previously available methods, and rBET (Model 1) trained on only one thin slice CT using data augmentation, on the test dataset of 22 head CT images. rBET+DA: rBET trained with data augmentation (DA). rBET: TS rBET trained on only one thin slice CT without data augmentation. ITK SS: ITK skull stripping filter; CT BET: Modified MRI brain extraction tool for CT head images. rBET: Our best CNN model (1).

Methods	Dice	Accuracy	Precision	Recall	Execution time(seconds)
ITK SS	0.949 ± 0.023	0.986 ± 0.007	0.959 ± 0.022	0.941 ± 0.032	~30–360 s
CT BET	0.995 ± 0.002	0.999 ± 0.001	0.994 ± 0.003	0.995 ± 0.002	~ 20 s
rBET	0.998 ± 0.001	1.000 ± 0.000	0.998 ± 0.001	0.999 ± 0.001	2 s
TS rBET	0.979 ± 0.005	0.994 ± 0.001	0.997 ± 0.002	0.963 ± 0.009	~2 s
rBET+DA	0.989 ± 0.003	0.997 ± 0.001	0.989 ± 0.004	0.989 ± 0.004	~2 s

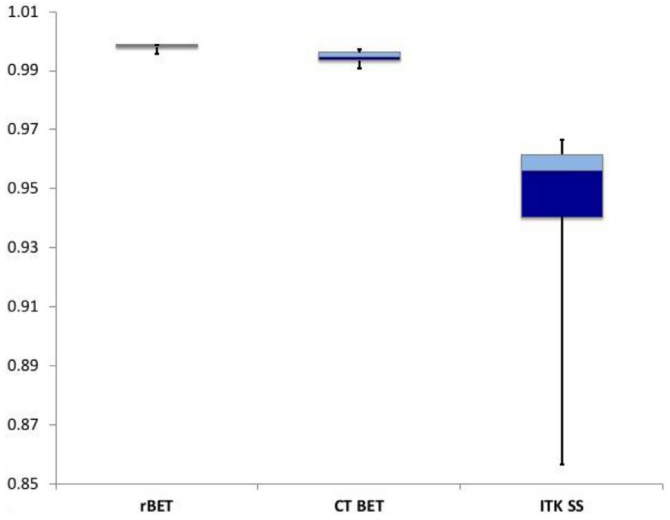

Fig. 5. The boxplot for the DICE performance of rBET, CT BET, and ITK skullstripping (ITK SS) across 22 CT head images. The bottom and top ends of whiskers represent minimum and maximum points.

Table 4
The performance (mean ± standard deviation) of CT BET and our CNN models on test dataset of 22 head CT images reconstructed with bone kernel. rBET1: Our best CNN model (1) trained with head CT images reconstructed with bone kernel only. rBET2: Our best CNN model (1) trained with head CT images reconstructed both with bone and soft-tissue kernels and tested on images reconstructed with bone kernel only. rBET3: The performance of rBET2 on head CT images reconstructed with soft-tissue kernel.

Model 1	Dice	Accuracy	Precision	Recall
CT BET	0.845 ± 0.031	0.952 ± 0.011	0.768 ± 0.045	0.941 ± 0.029
rBET1	0.995 ± 0.001	0.999 ± 0.000	0.993 ± 0.001	0.998 ± 0.001
rBET2	0.995 ± 0.001	0.999 ± 0.000	0.993 ± 0.001	0.998 ± 0.002
rBET3	0.996 ± 0.001	0.999 ± 0.000	0.993 ± 0.001	0.999 ± 0.001

CT head images reconstructed with bone kernel. The typical time to extract the brain was 2 s with GPU for our approach, which is much faster than other approaches. Fig. 6 shows absolute error between manual ground truth obtained from the Radiologist and our best CNN model, ITK filter, and CT BET methods for an example case. Figs. 7–9 shows Bland-Altman plots [3] for the agreement between two experts' manual volume segmentations and the agreement between experts 1 and 2 manual volume segmentation and CNN model, respectively.

We visually observed acceptable results for 3 post-surgical, 2 neonatal, and 52 hemorrhage cases. We visually confirmed that our brain extraction method based on CNN preserved the extra-axial hemorrhage within brain masks in all 52 cases.

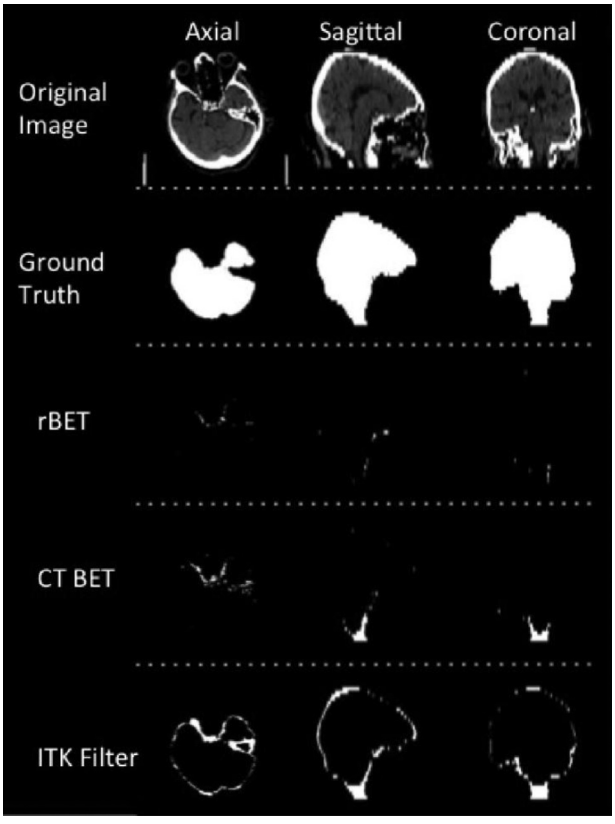

Fig. 6. Absolute error between ground truth and the results of our method (rBET), CT BET, and ITK skullstripping filter.

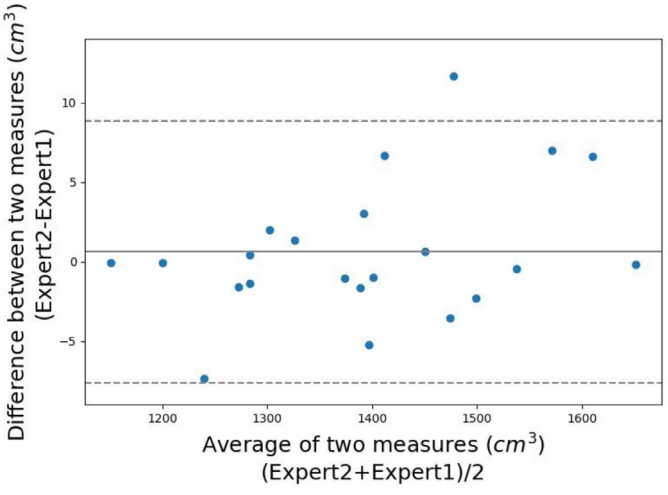

Fig. 7. Bland–Altman plot for inter-expert agreement for manual volume segmentation of two experts.

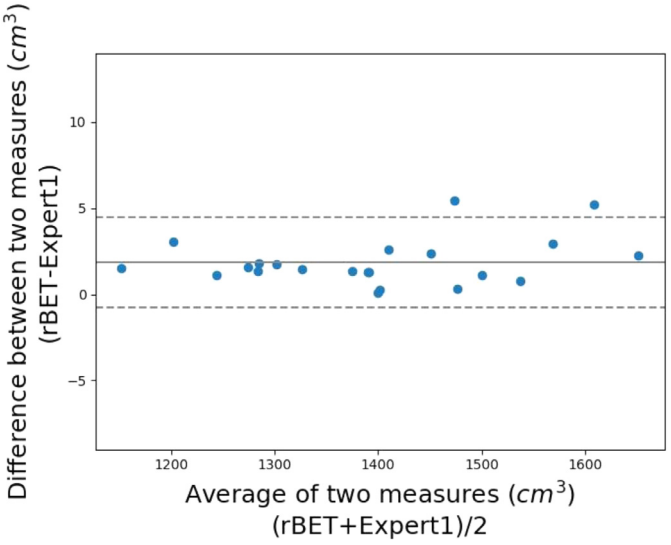

Fig. 8. Bland–Altman plot for the agreement of volume measurements between our best CNN model (1) and manual ground truth obtained from Expert 1.

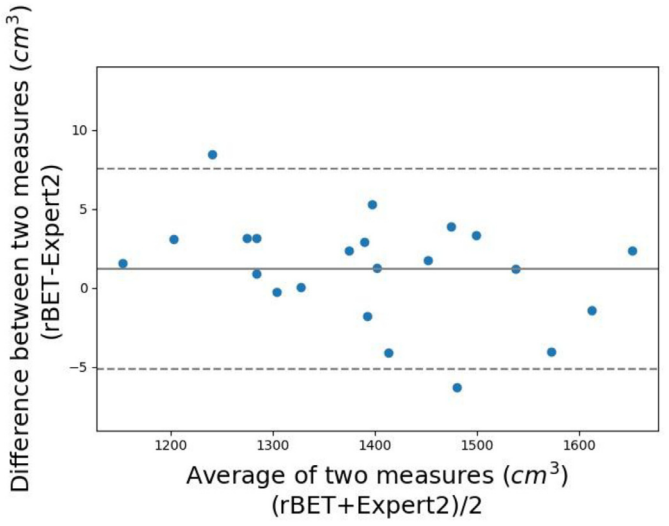

Fig. 9. Bland–Altman plot for the agreement of volume measurements between our best CNN model (1) and manual ground truth obtained from Expert 2.

4. Discussion

We present a robust and accurate brain extraction tool for head CT scans. To the best of our knowledge, this is the first study using deep fully convolutional neural networks to remove non-brain tissue from head CT images in an accurate, reliable, and time efficient way. It takes only about 2 s for our algorithm to remove non-brain tissue from head CT with 3 mm slices using an NVIDIA Titan X GPU, which is much faster than other approaches [5,16] for removing non-brain tissue from head CT images (2 s vs. 20 s-4 min). As seen in the results section, our method outperforms other publicly available and published methods for CT brain extraction.

The average performance of CT BET is very close to the performance of our method with higher standard deviation over our test dataset for the images reconstructed with soft tissue kernel. It does however show poor performance when used on the images reconstructed with the bone kernel whereas our method performs well on both reconstructions. This shows the generalization ability of our method as it works on CT images reconstructed with bone kernel that are quite noisy. In general, CT BET using the threshold and smoothing parameters suggested by Muschelli et al. [16] provides decent results. Even though CT BET provides acceptable results in head CT images reconstructed with soft tissue kernel but it occasionally includes small patches from the orbits. This was not observed in the results of our method. The performance of ITK skull stripping filter [5] showed the poorest average performance with the highest variability for this task. This could be due to the initialization of the level set algorithm with erroneous registration of an atlas image to patient images.

As shown in Table 1, CNN model 1 gave the best performance and modified versions of model 1 (Models 2 and 3) did not improve the performance of the network further. Including residual connections and batch normalization in model 3 dropped the performance. The 3D CNN model (model 5) showed poorer performance on the test dataset when compared to the 2D models. This is most likely because of down-sampling of input volumes to fit the GPU's memory. Having a GPU card with enough memory and using original volume as input to train 3D CNN model could further improve the performance of 3D model.

The variability in the performance of our method (CNN Model 1) over the 22 cases in the test dataset is very small, which shows the generalization ability of our method. A common challenge in developing segmentation methods is creating a high quality 'ground truth', and we note that many of the errors observed were along the margins with the skull base, where partial volume averaging makes the margins between brain and non-brain tissue inherently imprecise. As shown in Figs. 7–9, the variability in measuring brain volumes with our method is on the same order with the variability between manual volume measurements. This means that our method is as good as expert segmentations.

Data augmentation did not improve the performance of our best model (model 1) further on 22 test dataset but it did help to assure good generalization by training only one thin slice CT head image acquisition as shown in Table 3. Thin slice acquisition allowed us to resample the head CT volumes at different angles allowing us to create thousands of 2D slices for training CNN models. This drastically reduced the amount of data and tedious manual ground truth segmentations that is necessary for the training of CNN models. Including a few more thin-slice CT head images in the training and using aggressive data augmentation could potentially improve performance.

We did not include CT scans with extra-axial hemorrhage, substantial post-surgical changes or CT scans of neonates in the training dataset. We qualitatively observed acceptable segmentation results on post-op scans, neonatal head CT images, and CT scans with extra-axial hemorrhage but it is necessary to quantitatively assess the performance of CNN models on these cases. A limitation of this study is using only data from a single center and a single CT scanner. However, the weights of our current CNN model 1 could be used as a good initialization to transfer the learned knowledge and apply further training for tuning our fully CNN model for multicenter and multivendor data and the applications mentioned above.

5. Conclusion

Our approach based on fully convolutional neural networks extracts brain tissue from head CT images in an accurate, reliable, and computationally efficient manner. This will eliminate tedious and laborious manual segmentation of brain tissue from CT head images and provide a fast brain extraction tool for further processing or screening of brain diseases from non-contrast CT head images. Data augmentation on thin-slice head CT images could be used to improve generalization ability of fully CNN and reduce the amount of training dataset and tedious manual ground truth.

Please cite this article as: Z. Akkus, P. Kostandy and K.A. Philbrick et al., Robust brain extraction tool for CT head images, Neurocomputing, https://doi.org/10.1016/j.neucom.2018.12.085

Conflict of interest

None.

Acknowledgements

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

References

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, TensorFlow: A System For Large-Scale Machine Learning, OSDI., 2016.
[2] Z. Akkus, A. Galimzianova, A. Hoogi, D.L. Rubin, B.J. Erickson, Deep learning for brain MRI segmentation: state of the art and future directions, J. Digit. Imaging 30 (4) (2017) 449–459, doi:10.1007/s10278-017-9983-4.
[3] D.G. Altman, J.M. Bland, Measurement in medicine: the analysis of method comparison studies, J. R. Stat. Soc. Ser. D (Statistician) 32 (1983) 307–317.
[4] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a Deep convolutional encoder–decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 2481–2495.
[5] S. Bauer, T. Fejes, M. Reyes, A skull-stripping filter for ITK, Insight J. 2012 (2013) Jan–Dec. http://hdl.handle.net/10380/3353.
[6] M. Chawla, S. Sharma, J. Sivaswamy, L.T. Kishore, A method for automatic detection and classification of stroke from brain CT images, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, doi:10.1109/iembs.2009.5335289.
[7] Chollet, F. (2015) Keras, GitHub. https://github.com/fchollet/keras.
[8] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010, pp. 249–256.
[9] M.S. Holia, V.K. Thakar, Mutual information based image registration for MRI and CT SCAN brain images, in: Proceedings of the International Conference on Audio, Language and Image Processing, 2012, doi:10.1109/icalip.2012.6376590.
[10] J.E. Iglesias, C.-Y. Liu, P.M. Thompson, Z. Tu, Robust brain extraction across datasets and comparison with publicly available methods, IEEE Trans. Med. Imaging 30 (2011) 1617–1634.
[11] M. Jenkinson, C.F. Beckmann, T.E.J. Behrens, M.W. Woolrich, S.M. Smith, FSL, Neuroimage 62 (2012) 782–790.
[12] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization - arXiv preprint arXiv: 1412.6980, 2014.
[13] Y. Li, Q. Hu, J. Wu, Z. Chen, A hybrid approach to detection of brain hemorrhage candidates from clinical head CT scans, in: Proceedings of the Sixth International Conference on Fuzzy Systems and Knowledge Discovery, 2009, pp. 361–365.
[14] J.G. Mandell, J.W. Langelaan, A.G. Webb, S.J. Schiff, Volumetric brain analysis in neurosurgery: part 1. Particle filter segmentation of brain and cerebrospinal fluid growth dynamics from MRI and CT images, J. Neurosurg. Pediatr. 15 (2015) 113–124.
[15] F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: fully convolutional neural networks for volumetric medical image segmentation, in: Proceedings of the Fourth International Conference on 3D Vision (3DV), 2016, doi:10.1109/3dv.2016.79.
[16] J. Muschelli, N.L. Ullman, W.A. Mould, P. Vespa, D.F. Hanley, C.M. Crainiceanu, Validated automatic brain extraction of head CT images, Neuroimage 114 (2015) 379–385.
[17] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.
[18] R. Peter, P. Korfiatis, D. Blezek, A. Oscar Beitia, I. Stepan-Buksakowska, D. Horinek, K.D. Flemming, B.J. Erickson, A quantitative symmetry-based analysis of hyperacute ischemic stroke lesions in noncontrast computed tomography, Med. Phys. 44 (2017) 192–199.
[19] R. Peter, P. Korfiatis, D. Blezek, A. Oscar Beitia, I. Stepan-Buksakowska, D. Horinek, K.D. Flemming, B.J. Erickson, A quantitative symmetry-based analysis of hyperacute ischemic stroke lesions in noncontrast computed tomography, Med. Phys. 44 (2017) 192–199.
[20] O. Ronneberger, P. Fischer, T. Brox, N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi, U-Net: convolutional networks for biomedical image segmentation, in: Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI, Springer International Publishing, Cham, 2015, pp. 234–241. Lecture Notes in Computer Science.
[21] C. Rorden, L. Bonilha, J. Fridriksson, B. Bender, H.-O. Karnath, Age-specific CT and MRI templates for spatial normalization, Neuroimage 61 (2012) 957–965.
[22] C. Rorden, M. Brett, Stereotaxic display of brain lesions, Behav. Neurol. 12 (2000) 191–200.
[23] W. Schroeder, L. Ng, J. Cates, The ITK Software Guide, The Insight Consortium, 2003.
[24] S.M. Smith, Fast robust automated brain extraction, Hum. Brain Mapp 17 (2002) 143–155.
[25] M. Sun, R. Hu, H. Yu, B. Zhao, H. Ren, Intracranial hemorrhage detection by 3D voxel segmentation on brain CT images, in: Proceedings of the International Conference on Wireless Communications & Signal Processing (WCSP), 2015, doi:10.1109/wcsp.2015.7341238.
[26] H.-L. Tong, M.F.A. Fauzi, S.-C. Haw, Automated hemorrhage slices detection for CT brain images, in: Proceedings of the International Visual Informatics Conference, Berlin, Heidelb erg, Springer, 2011, pp. 268–279.
[27] P.A. van den Elsen, M.A. Viergever, n.d. Automated CT and MR brain image registration using geometrical feature correlation, in: Proceedings of the IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference. doi:10.1109/nssmic.1993.373608.
[28] Y. Wang, J. Nie, P.-T. Yap, G. Li, F. Shi, X. Geng, L. Guo, D. Shen, Alzheimer's Disease Neuroimaging Initiative, Knowledge-guided robust MRI brain extraction for diverse large-scale neuroimaging studies on humans and non-human primates, PLoS One 9 (2014) e77810.
[29] Y. Xia, J. Wang, S. Eberl, M. Fulham, D.D. Feng, Brain tissue segmentation in PET-CT images using probabilistic atlas and variational Bayes inference, Conf. Proc. IEEE Eng. Med. Biol. Soc. 2011 (2011) 7969–7972.
[30] J. Yang, L. Lu, S. Ma, W. Tan, D. Zhao, An automatic method to extract brain tissue in CT data, in: Proceedings of the Eighth International Conference on Information Technology in Medicine and Education (ITME), 2016, pp. 70–72.
[31] P.A. Yushkevich, J. Piven, H.C. Hazlett, R.G. Smith, S. Ho, J.C. Gee, G. Gerig, User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability, Neuroimage 31 (2006) 1116–1128.

Zeynettin Akkus received his M.Sc. degree in medical imaging from the Medical Engineering Department of the School of Health and Technology, KTH, the Royal Institute of Technology, Stockholm, Sweden, in 2009. He worked as a visiting researcher in the Ultrasound Research Group, Department of Medical Physics, University Hospitals of Leicester, Royal Infirmary, in 2009. He received his PhD on image analysis of *contrast-enhanced ultrasound carotid plaque imaging* from Biomedical Engineering Department, Erasmus Medical Center, Rotterdam, The Netherlands. *He developed a set of image processing tools and software packages for medical image analysis.* His research interests are biomedical image and signal processing, angiogenesis imaging and quantification, deep learning in medical image analysis, and decision support systems. He received IEEE Best Paper Award in 2013 and several poster awards. He was also a member of the team who won the NVIDIA Global Impact award in 2017 for the prediction of genomic biomarker status from MR images using deep learning.

Petro Kostandy is a radiologist at Mayo Clinic, Rochester, MN. He received his degree in Medicine from Weill Cornell Medical College in Qatar and completed his residency at SUNY upstate Medical University. He also completed informatics radiology fellowship at Mayo Clinic. He is currently clinical fellow in Radiology at Mayo Clinic. His research interests are clinical decision support system, neuroradiology, and AI in radiology.

Kenneth A. Philbrick received his PhD from Oregon State University in 2013. Kenneth's research interests are broadly in utilizing deep learning to perform medical imaging inference, understandable AI, and deep learning tool development. Kenneth has developed, RIL-Contour, a Python based medical imaging data annotation tool designed to accelerate dataset annotation. Kenneth has developed the "SAM" deep learning layer visualization technique to visualize activations of intermediate network layers; described in the paper, *What Does Deep Learning See: Insights from a Classifier Trained to Predict Contrast Enhancement Phase from CT Images.* AJR Dec. 2018.

Bradley J. Erickson received his MD and Ph.D. degrees from Mayo Medical & Graduate Schools and then did his residency in diagnostic radiology and Neuroradiology fellowship at Mayo Clinic. He went on staff at Mayo Clinic, and was heavily involved in implementing a filmless department as well as being the Vice Chair for IT. More recently, he has refocused on imaging informatics research, receiving NIH grants for brain cancer, multiple sclerosis, and polycystic kidney disease. He is particularly focused on the application of deep learning to medical images, which is the emerging field of radiogenomics. He was the founding Chair of the Division of Imaging Informatics, and has been the Associate Chair for Research in Radiology.